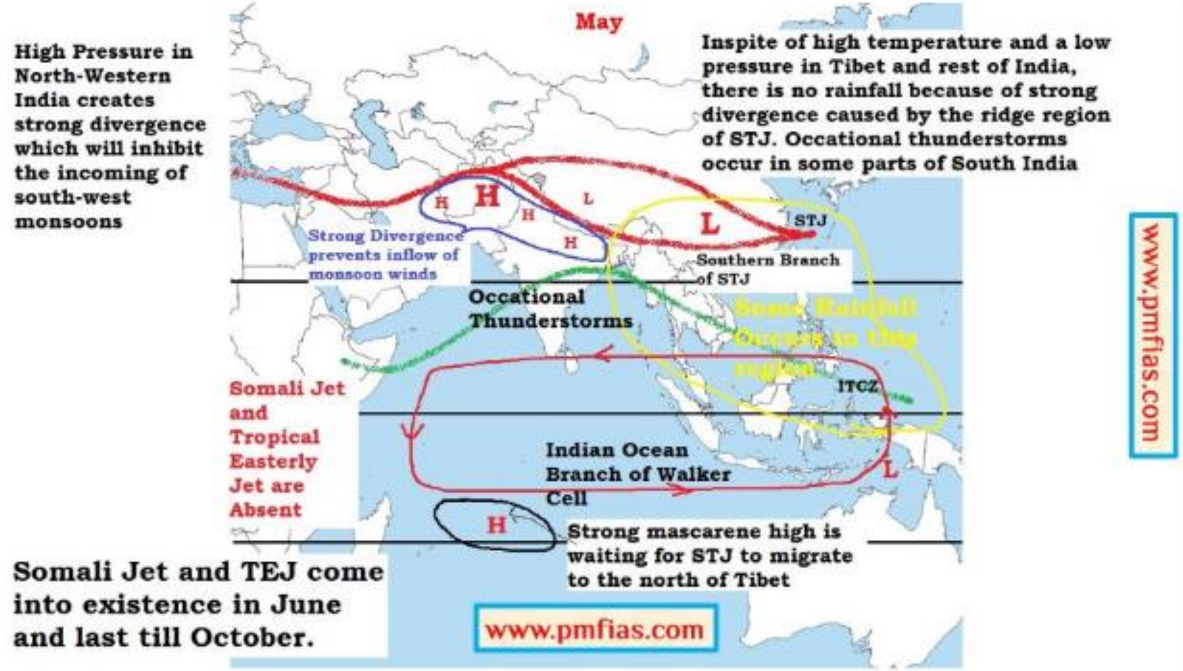
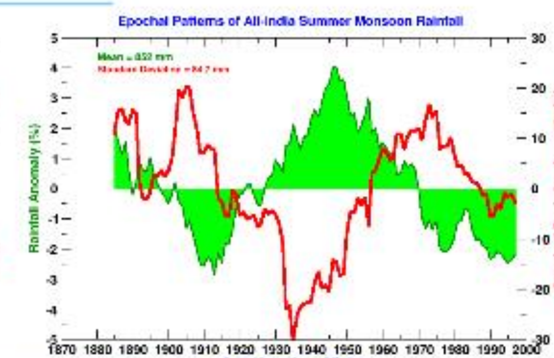
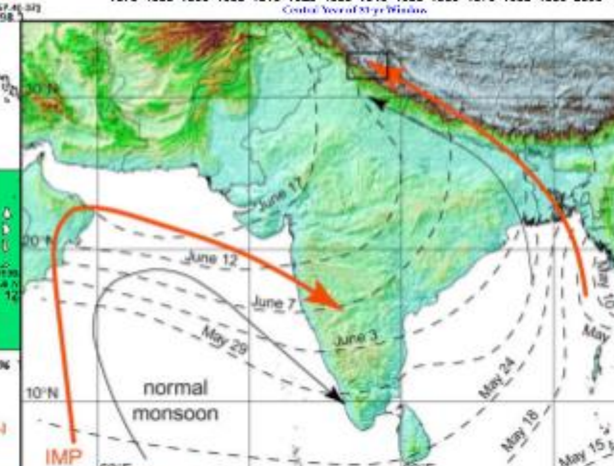
Indian Monsoon through the lens of a Discrete Random Field

Adway Mitra
International Center for Theoretical Sciences


High Pressure in North-Western India creates strong divergence which will inhibit the incoming of south-west monsoons
May
Inspite of high temperature and a low pressure in Tibet and rest of India, there is no rainfall because of strong divergence caused by the ridge region of STJ. Occational thunderstorms occur in some parts of South India
H
L
STJ
Strong Divergence prevents inflow of monsoon winds
Southern Branch of STJ
Occational Thunderstorms
Some Rainfall Occurs in this region
ITCZ
Somali Jet and Tropical Easterly Jet are Absent
Indian Ocean Branch of Walker Cell
Strong mascarene high is waiting for STJ to migrate to the north of Tibet
Somali Jet and TEJ come into existence in June and last till October.
www.pmfias.com
www.pmfias.com

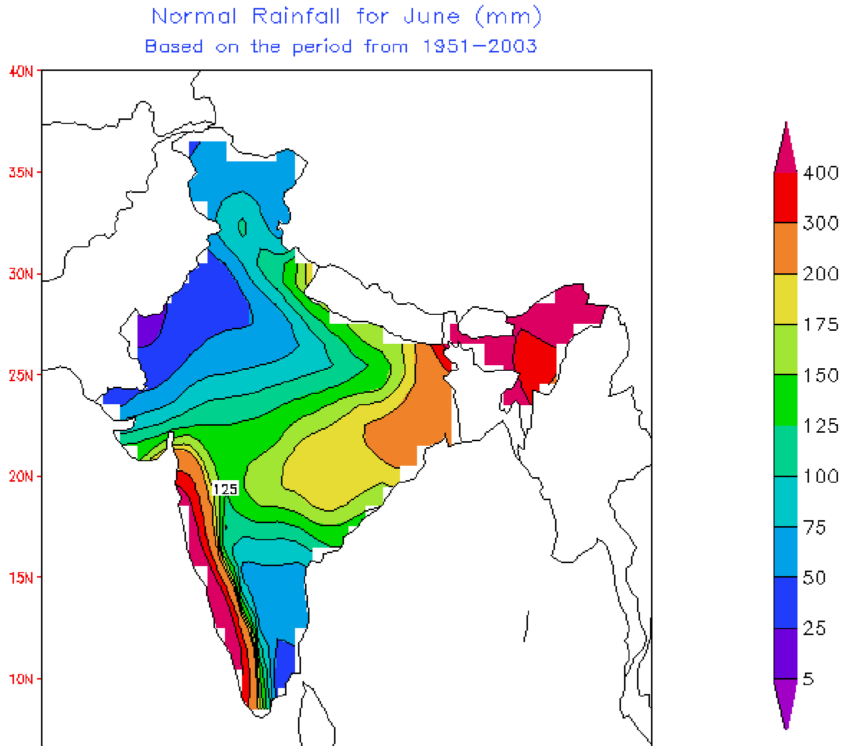

INDIA METEOROLOGICAL DEPARTMENT
जल मौसम विज्ञान प्रभाग
HYDROMET DIVISION, NEW DELHI
RAINFALL (mm.) FOR THE PERIOD
01.06.2014 TO 30.08.2014
PAKISTAN
CHINA
TIBET
NEPAL
BHUTAN
BAY OF BENGAL
ARABIAN SEA
INDIAN OCEAN
SRI LANKA
CATEGORYWISE NO. OF SUBDIVISIONS
LEGEND: EXCESS (+20% OR MORE) NORMAL (+19% TO -19%) DEFICIENT (-20% SCANTY (-60% TO -99%) NO RAIN (-100%) NO DATA
NOTES:
(a) Rainfall figures are based on operational data.
(b) Small figures indicate actual rainfall (mm.), while bold figures indicate Normal rainfall (mm.) Percentage Departures of Rainfall are shown in Brackets.

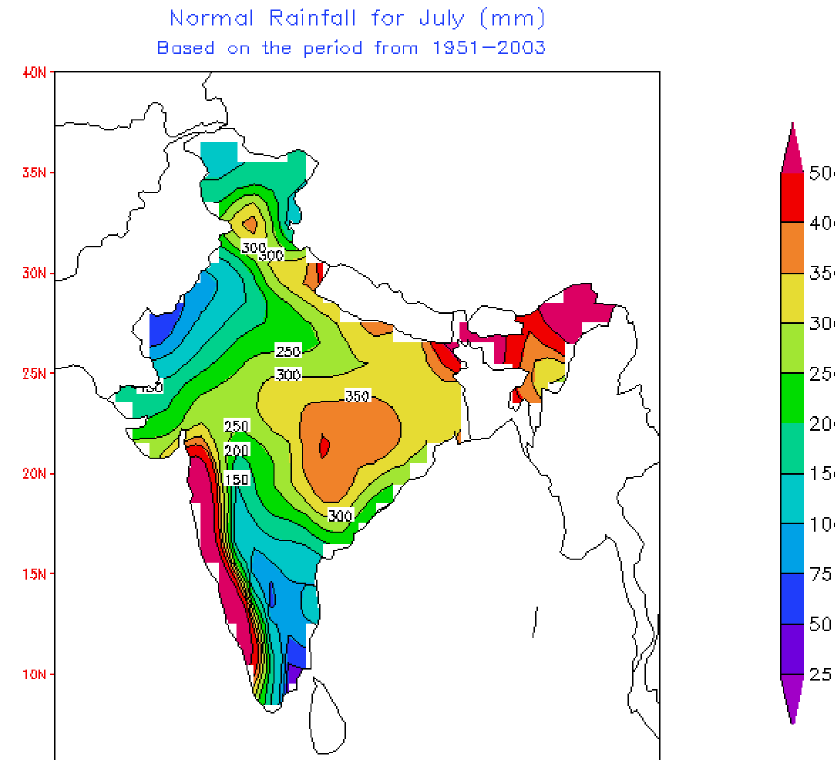

Epochal Patterns of All-India Summer Monsoon Rainfall
Rainfall Anomaly (%)
Standard Deviation Anomaly (%)
1870 1880 1890 1900 1910 1920 1930 1940 1950 1960 1970 1980 1990 2000

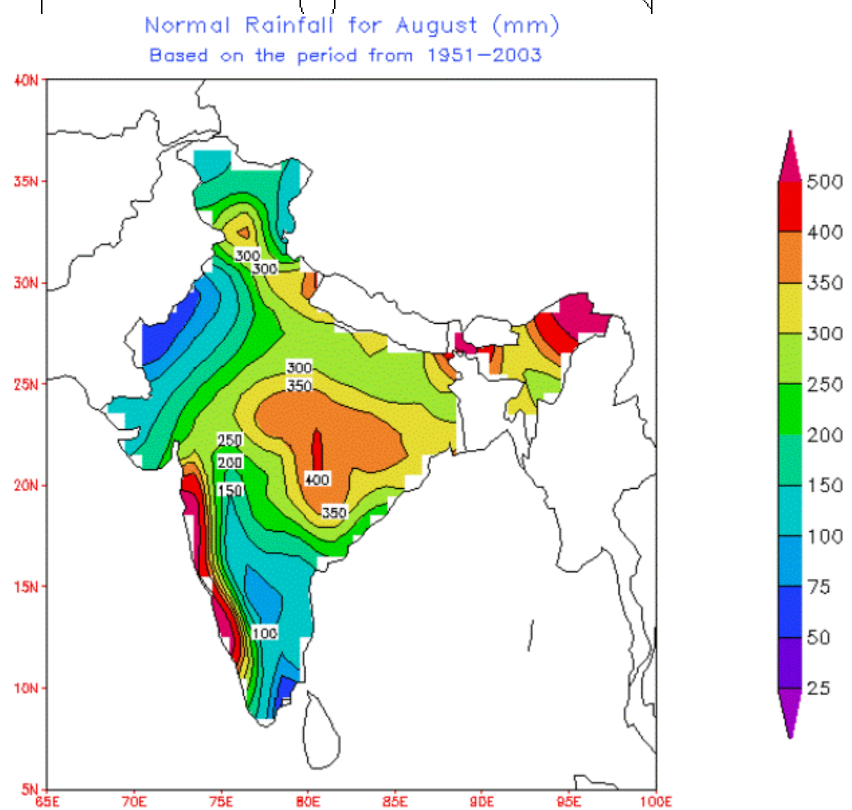

normal monsoon
IMP
June 17
June 12
June 7
June 3
May 29
May 24
May 18
May 15
May 10
30°N
20°N
10°N
60°E
80°E

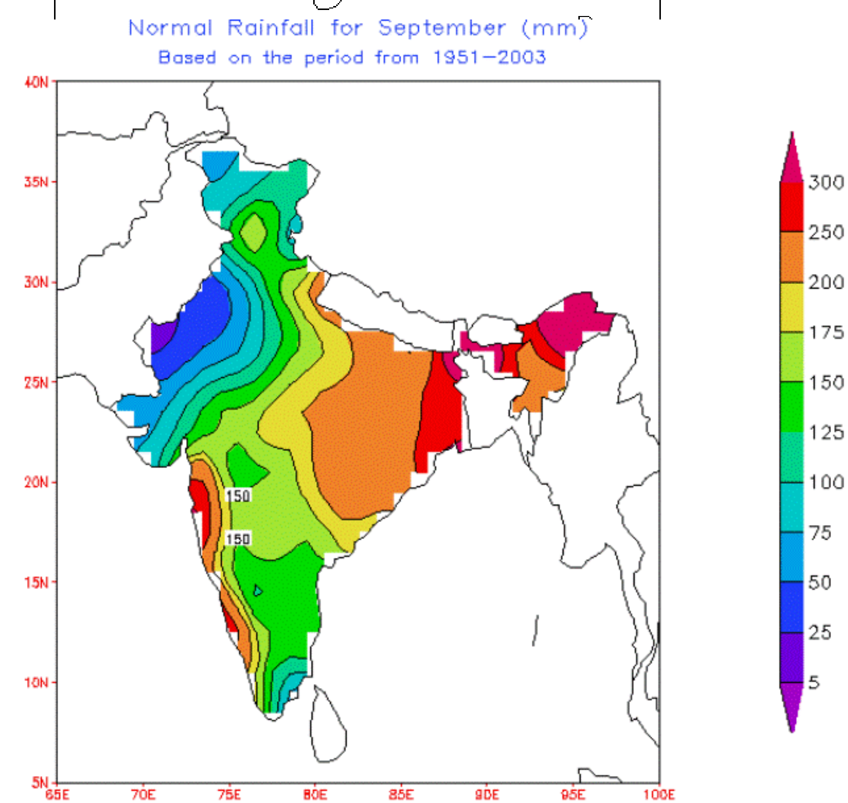
Notations

- S locations (357 grid-boxes all over India, 100Km-100Km)
- T time-steps (Each day in June-September, 2000-2007)
- X(s,t) : rainfall volume at location s on day t
- X(t): rainfall vector on day t
- Y(t): total rainfall on day t

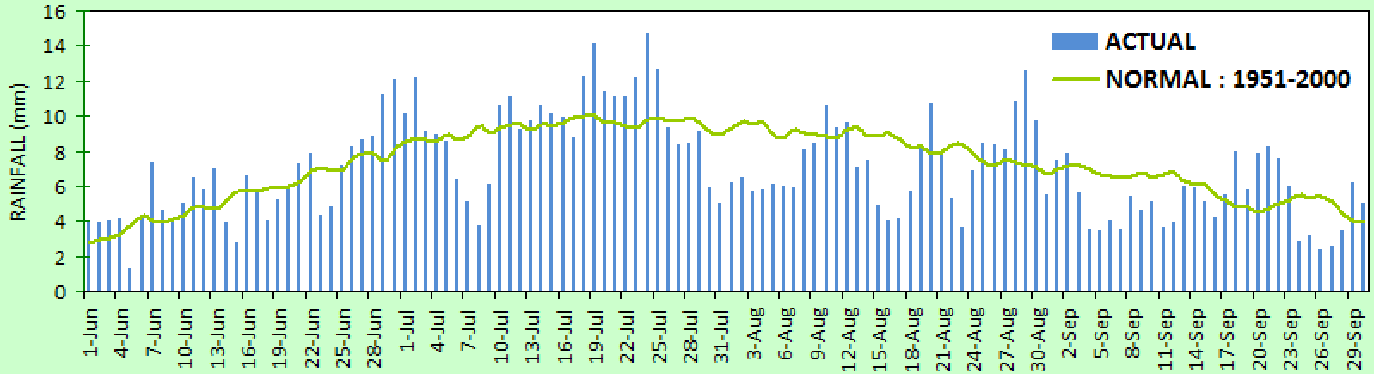

Normal Rainfall for June (mm)
Based on the period from 1951–2003
40N
35N
30N
25N
20N
15N
10N
125
400
300
200
175
150
125
100
75
50
25
5


Normal Rainfall for July (mm)
Based on the period from 1951–2003
40N
35N
30N
25N
20N
15N
10N
300
300
250
300
350
250
200
150
300
500
400
350
300
250
200
150
100
75
50
25

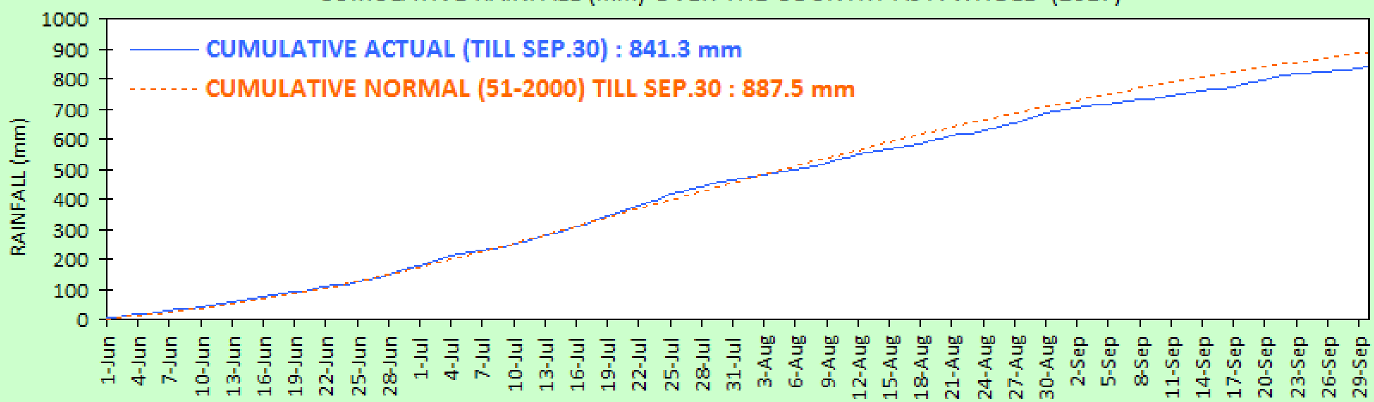

Normal Rainfall for August (mm)
Based on the period from 1951–2003
40N
35N
30N
25N
20N
15N
10N
5N
65E
70E
75E
80E
85E
90E
95E
100E
300
300
300
350
250
200
150
400
350
100
500
400
350
300
250
200
150
100
75
50
25


Normal Rainfall for September (mm)
Based on the period from 1951–2003
40N
35N
30N
25N
20N
15N
10N
5N
65E
70E
75E
80E
85E
90E
95E
100E
150
150
300
250
200
175
150
125
100
75
50
25
5

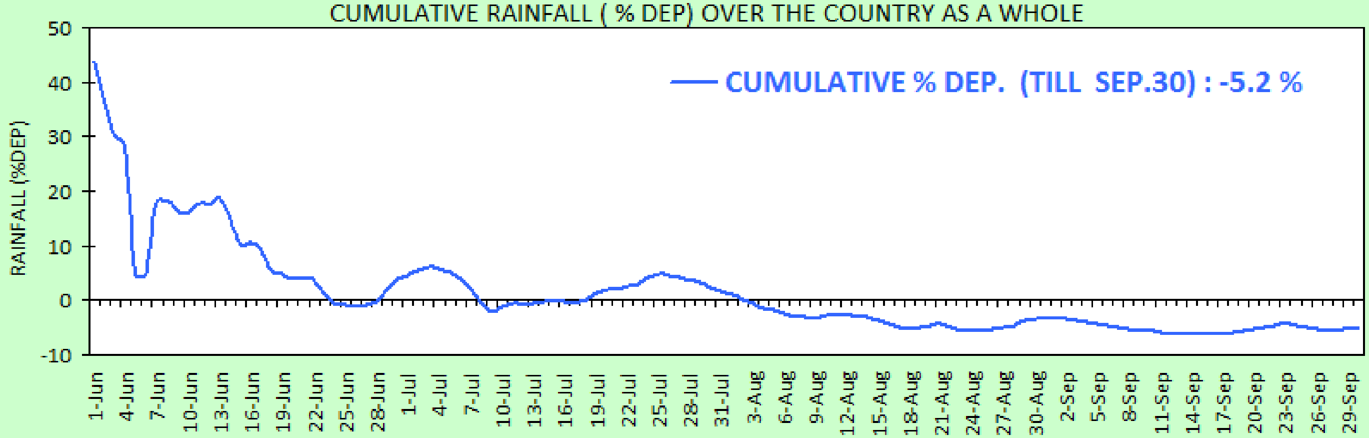
DAILY MEAN RAINFALL (mm) OVER THE COUNTRY AS A WHOLE (2017)



CUMULATIVE RAINFALL (mm) OVER THE COUNTRY AS A WHOLE (2017)

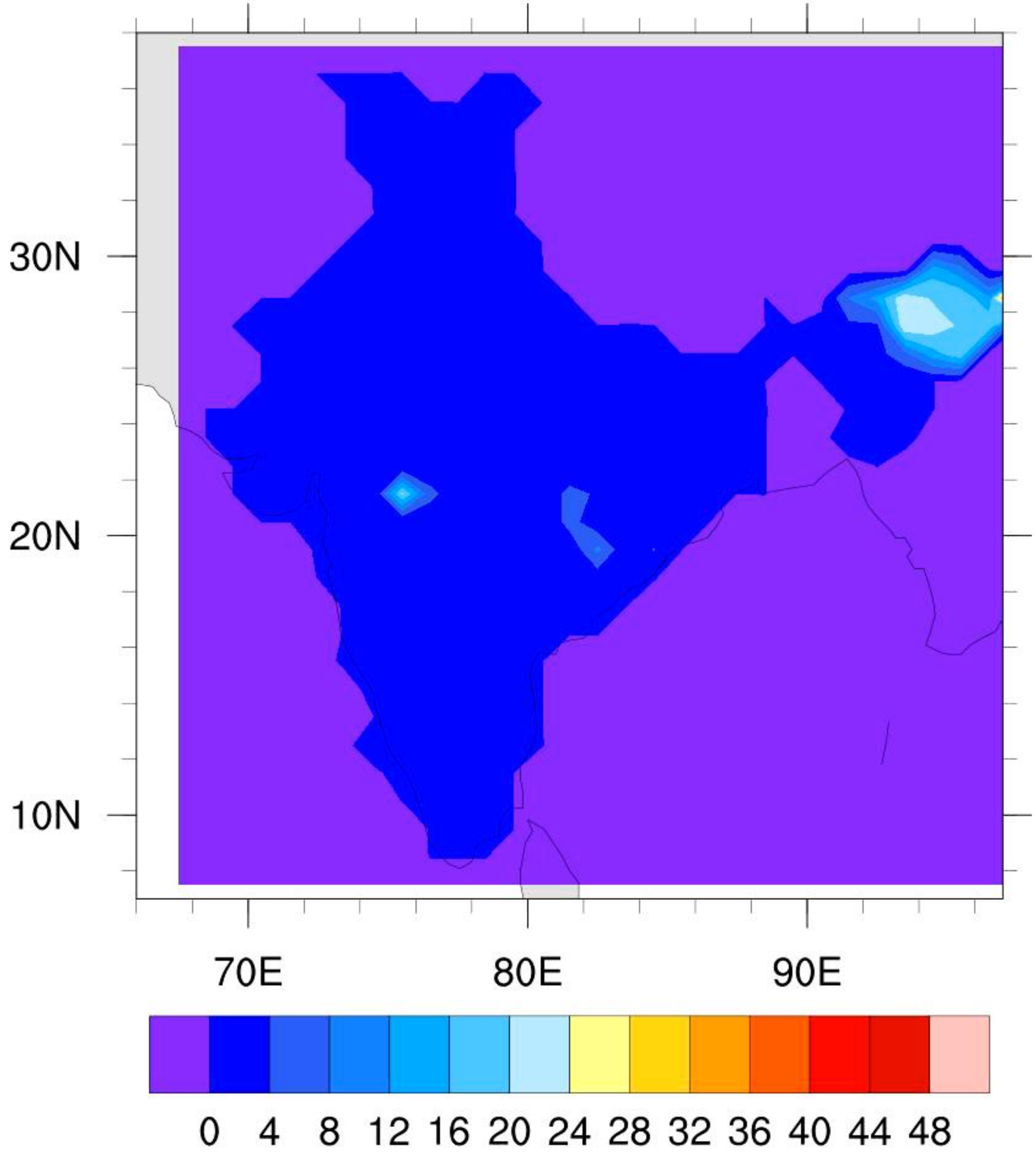


CUMULATIVE RAINFALL (% DEP) OVER THE COUNTRY AS A WHOLE



(वास्तविक समय के आंकड़ो पर आधारित) (Based on realtime data)

Is there a method in the madness?



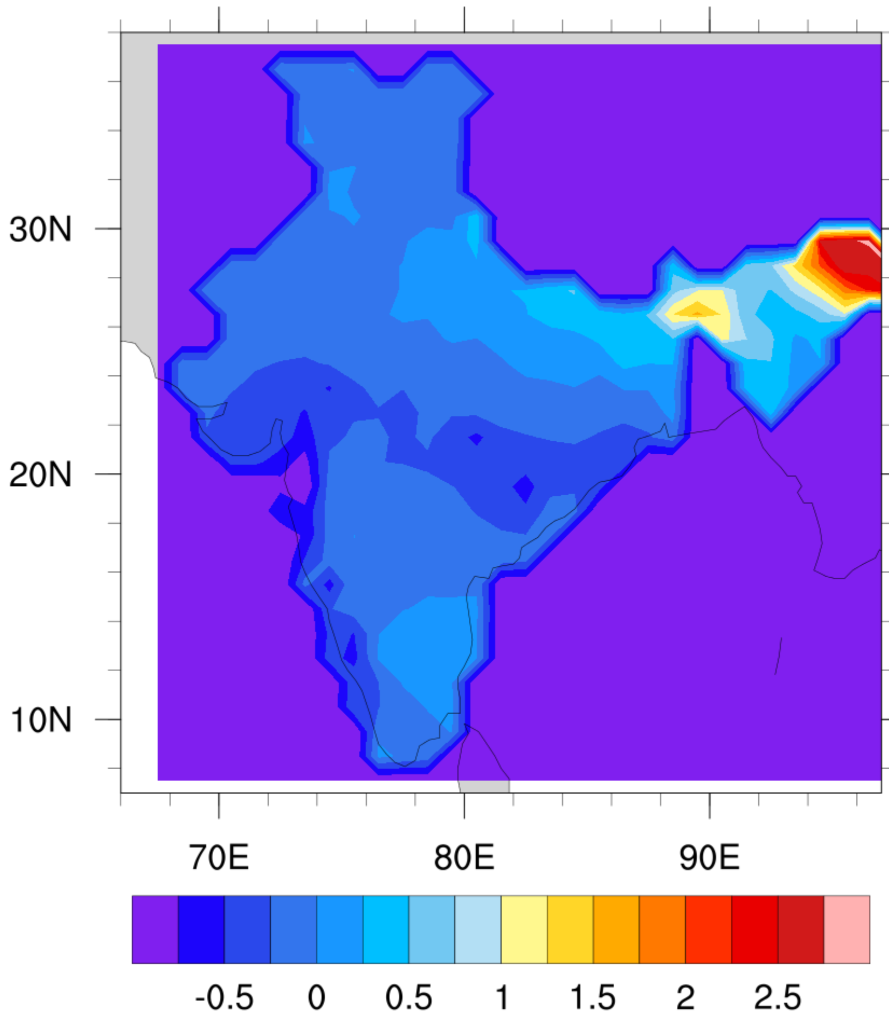
Aim of the Work

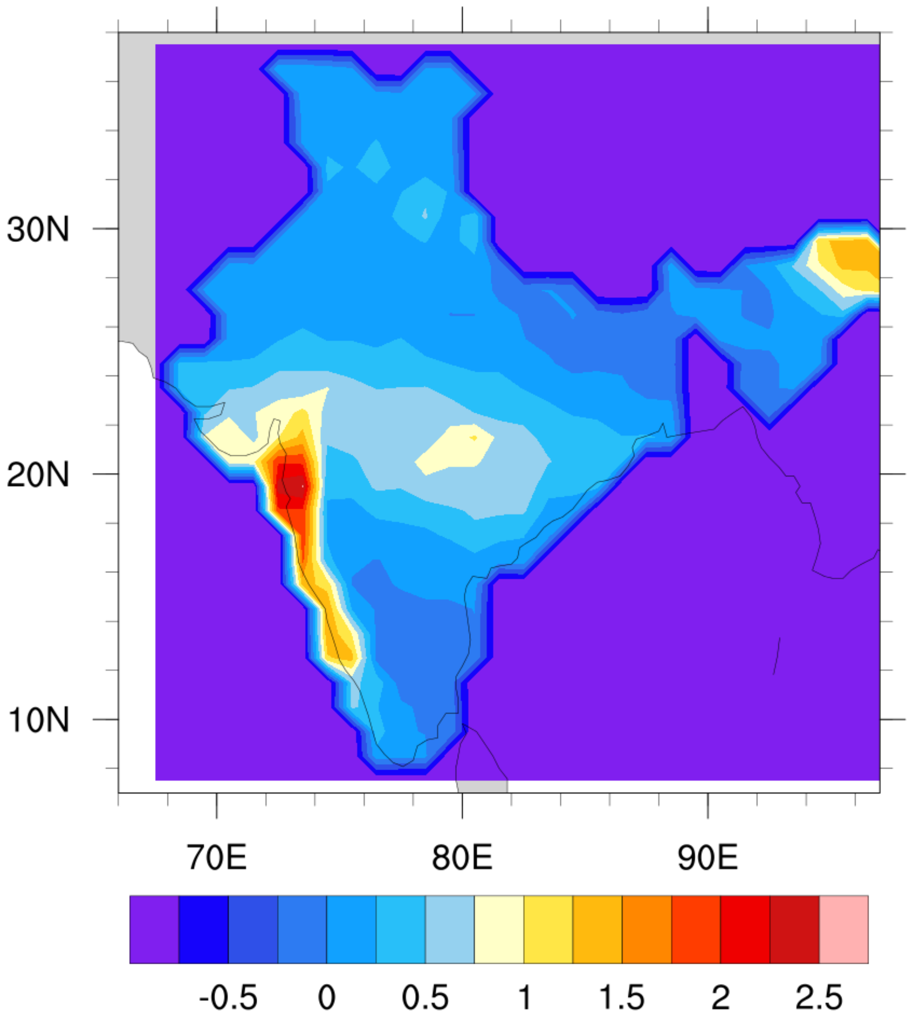
- Model for intra-seasonal spatio-temporal oscillations of monsoon
- Identify a set of “spatial patterns” of daily rainfall
- Represent each day’s rainfall distribution vector using these patterns
- Study transitions from one pattern to another
- The model should be general across all years

Empirical Orthogonal Function

- Usual approach: Empirical Orthogonal Function analysis
- Each eigenvector of sample covariance matrix represents a spatial pattern
- Each day’s rainfall vector is a linear combination of these spatial patterns
- Only first few “patterns” are significant

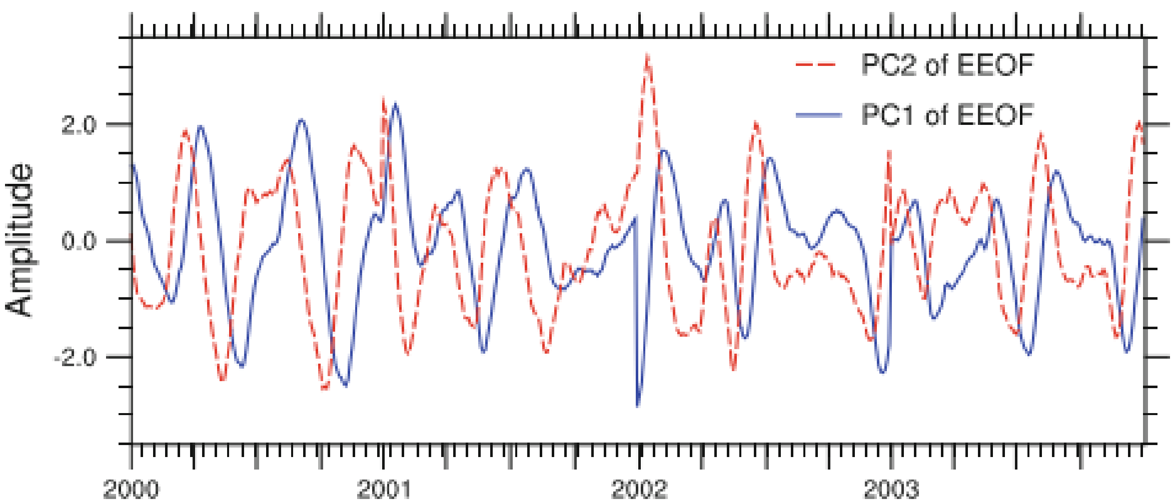
Empirical Orthogonal Functions

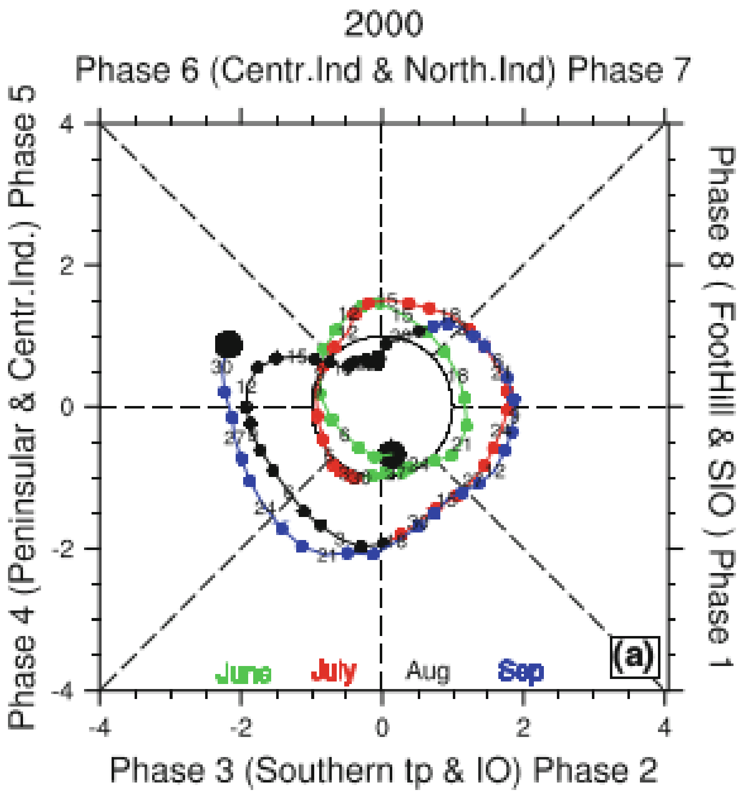




Monsoon Intra-seasonal Oscillations

- Suhas et al, Climate Dynamics, 2012





Drawbacks of EOF

- Lack of interpretability

  - Each vector a linear combination of many EOFs

  - EOFs are orthogonal to each other

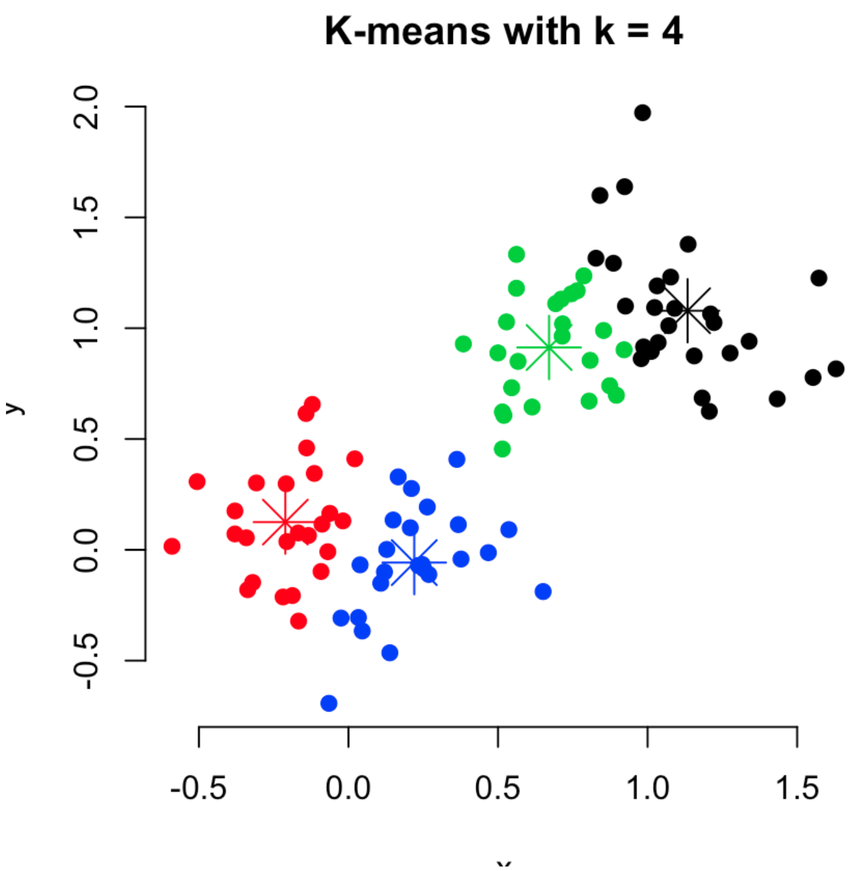
  - EOFs contain negative values

  - Coefficients may be negative
- Lack of Spatial Coherence

Clustering-based Approach

- Direct clustering of data vectors
- K-means, Spectral Clustering
- Each “cluster center” represents a spatial pattern
- Each data vector can be represented by one spatial pattern

K-means Clustering



number of clusters
number of cases
centroid for cluster *j*
case *i*

objective function ← $$J = \sum_{j=1}^{k} \sum_{i=1}^{n} \left\| x_i^{(j)} - c_j \right\|^2$$

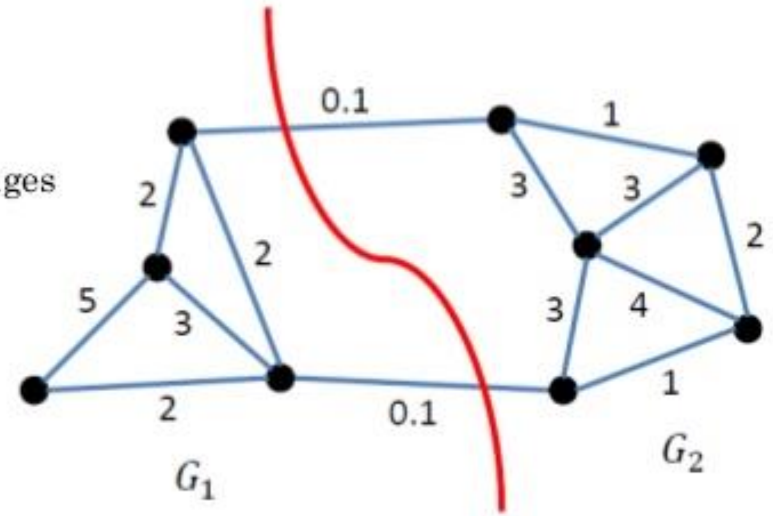
Distance function

Spectral Clustering

MINIMUM CUT

Sum of the weights to cut edges

$$\mathrm{cut}(G_1, G_2) = \sum_{i \in G_1, j \in G_2} w_{ij}$$



$$\min \sum_{i \in G_1, j \in G_2} w_{ij} \iff \min \boldsymbol{x}^{\mathrm{T}}(\boldsymbol{D} - \boldsymbol{W})\boldsymbol{x}$$

But, favor for small and isolated clusters

Spatial Patterns by K-means Clustering

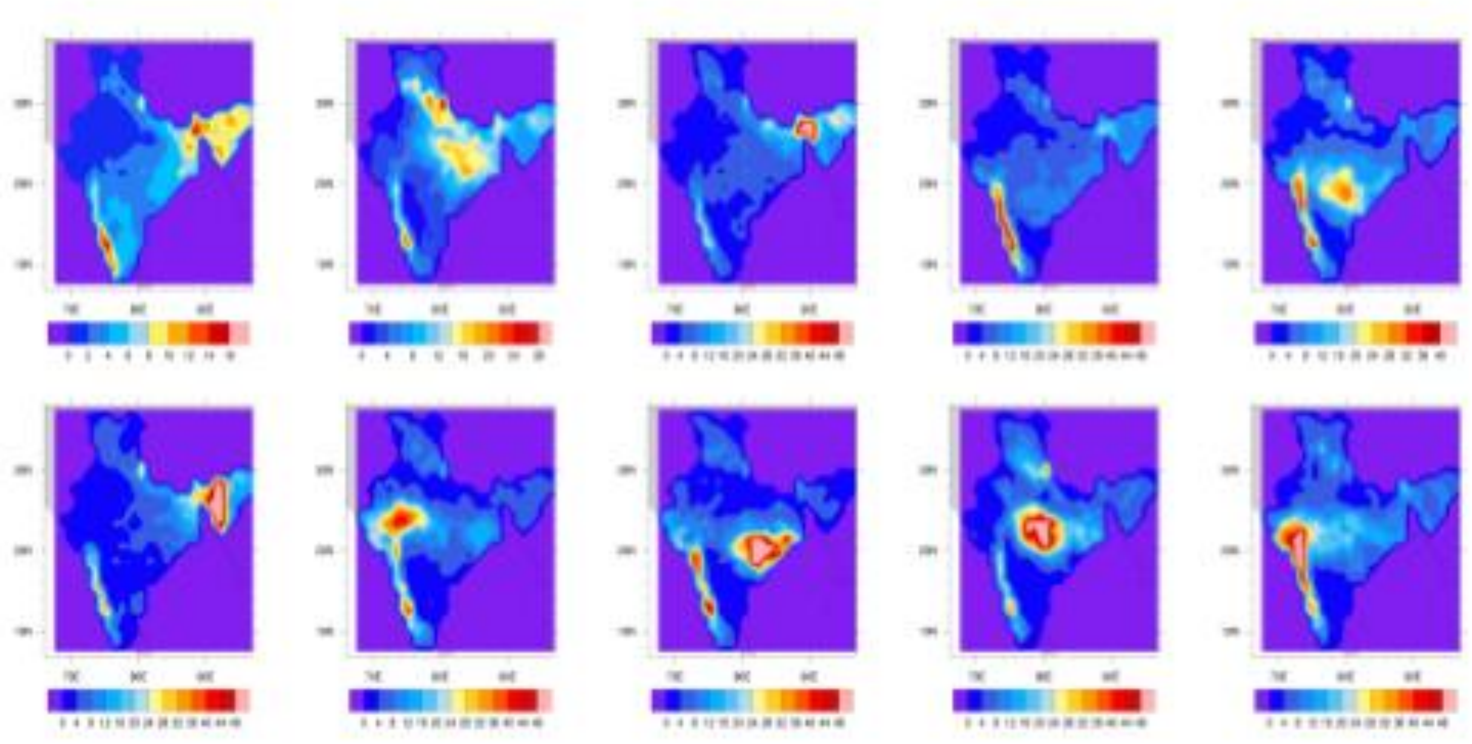


Figure 4: Canonical Rainfall Patterns (CRP) corresponding to 10 prominent clusters found by K-means

Drawbacks of Clustering

- The cluster “centers” rarely capture spatio-temporal properties
- How many clusters should be formed?
- Issues with interpretability remain
- Patterns not robust to the time-period

Discrete Representation

- A discrete representation of daily rainfall and spatial patterns may be more interpretable
- X(s,t): rainfall volume at location s on day t
- Z(s,t): binary representation of X(s,t)
- High X(s,t) -> Z(s,t) =1
- Low X(s,t) -> Z(s,t) =2


X(1,1)
Z(1,1)
X(2,1)
Z(2,1)
X(3,1)
Z(3,1)
X(S,1)
Z(S,1)
X(1,2)
Z(1,2)
X(2,2)
Z(2,2)
X(3,2)
Z(3,2)
X(S,2)
Z(S,2)
X(1,3)
Z(1,3)
X(2,3)
Z(2,3)
X(3,3)
Z(3,3)
X(S,3)
Z(S,3)
X(1,D)
Z(1,D)
X(2,D)
Z(2,D)
X(3,D)
Z(3,D)
X(S,D)
Z(S,D)

Climatic Interpretation

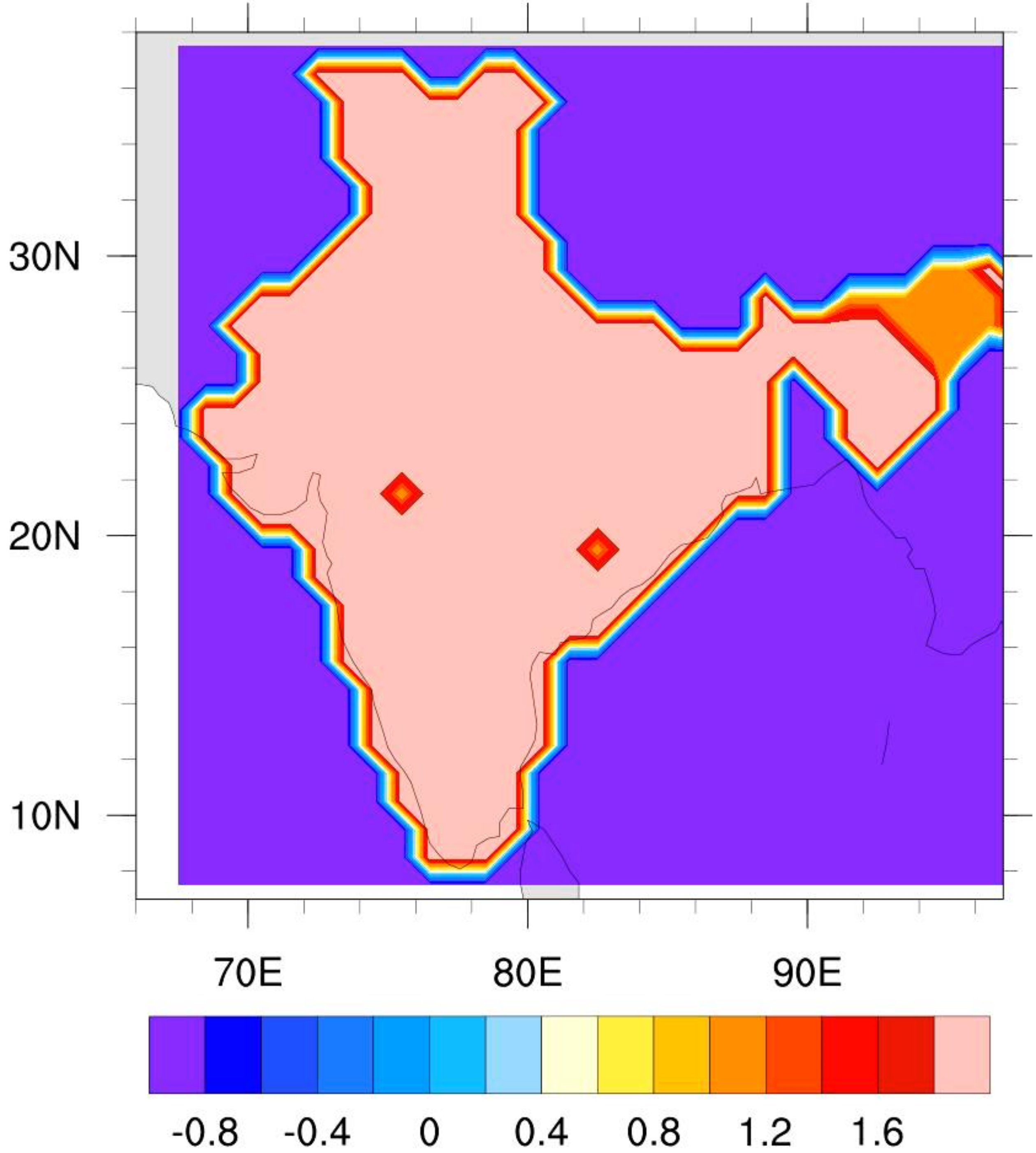
- Rainfall is the manifestation of underlying climatic processes
- Rainfall volumes may be localized but underlying processes are usually spatio-temporally extended
- Z-variables encode the underlying processes – are the climatic conditions conducive to rainfall?
- Z-variables should be spatio-temporally coherent

Threshold-based approach

- Put a threshold on X(s,t)
- Mean rainfall at location s across time-domain
- Problems remain:

  1) Spatio-temporal coherence of Z not guaranteed

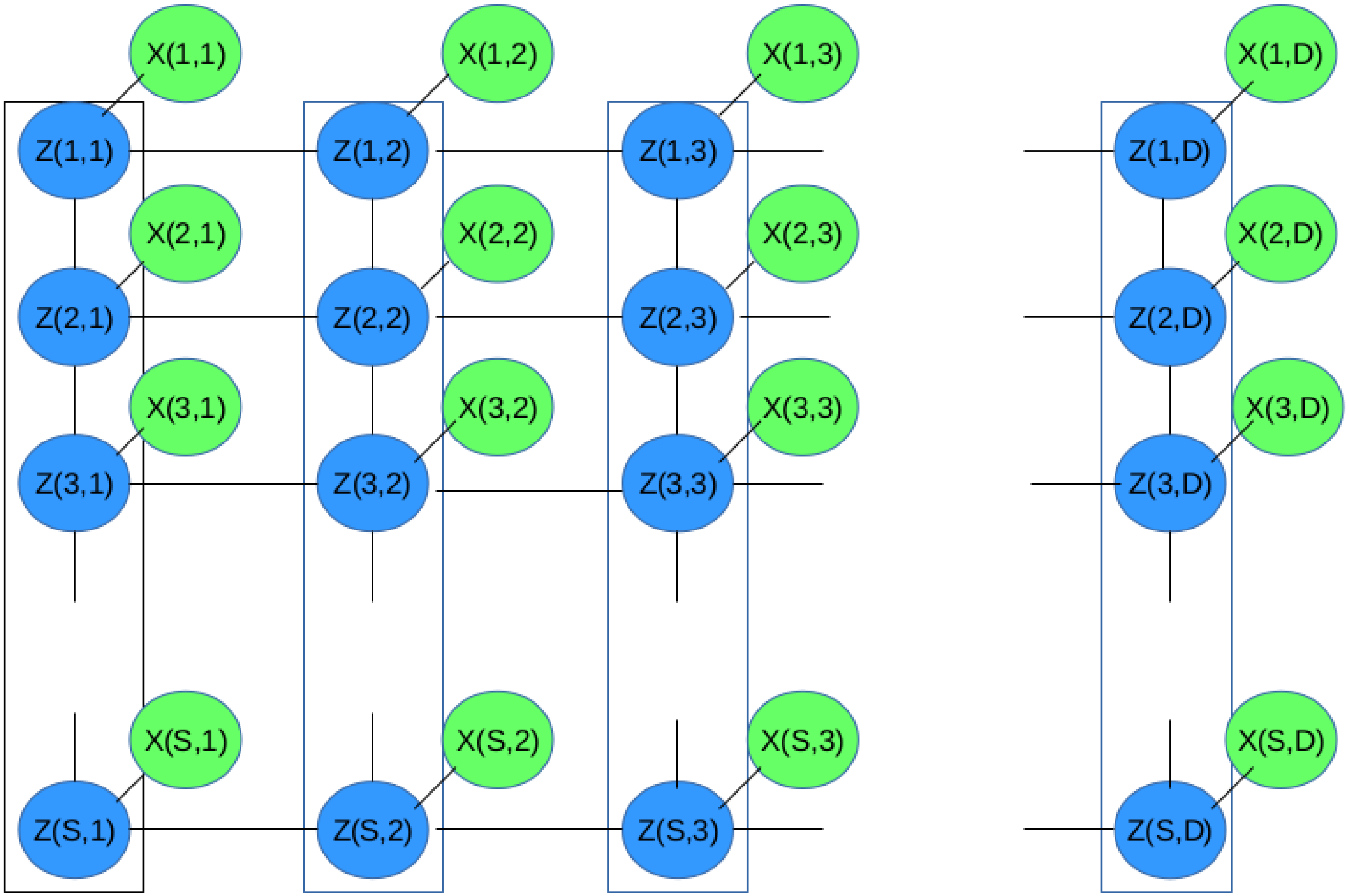
  2) How to get spatial patterns?



30N
20N
10N
70E
80E
90E
-0.8
-0.4
0
0.4
0.8
1.2
1.6

Markov Random Field

- Probabilistic State Assignment eliminates need of hard thresholds
- Spatio-temporal Coherence of Z-variables guaranteed
- Z,X considered random variables
- We consider a spatio-temporal graph of Z,X variables
- Z-variables connected based on spatio-temporal neighborhood


X(1,1)
X(1,2)
X(1,3)
X(1,D)
Z(1,1)
Z(1,2)
Z(1,3)
Z(1,D)
X(2,1)
X(2,2)
X(2,3)
X(2,D)
Z(2,1)
Z(2,2)
Z(2,3)
Z(2,D)
X(3,1)
X(3,2)
X(3,3)
X(3,D)
Z(3,1)
Z(3,2)
Z(3,3)
Z(3,D)
X(S,1)
X(S,2)
X(S,3)
X(S,D)
Z(S,1)
Z(S,2)
Z(S,3)
Z(S,D)

Markov Random Field Model

- Spatial Edges between Z(s,t) and Z(s’,t)
- Temporal Edges between Z(s,t) and Z(s,t’)
- Data Edges between Z(s,t) and X(s,t)
- Potential Functions Ψ on each edge

- Joint distribution of all Z and X variables:
- product of all edge potential functions!

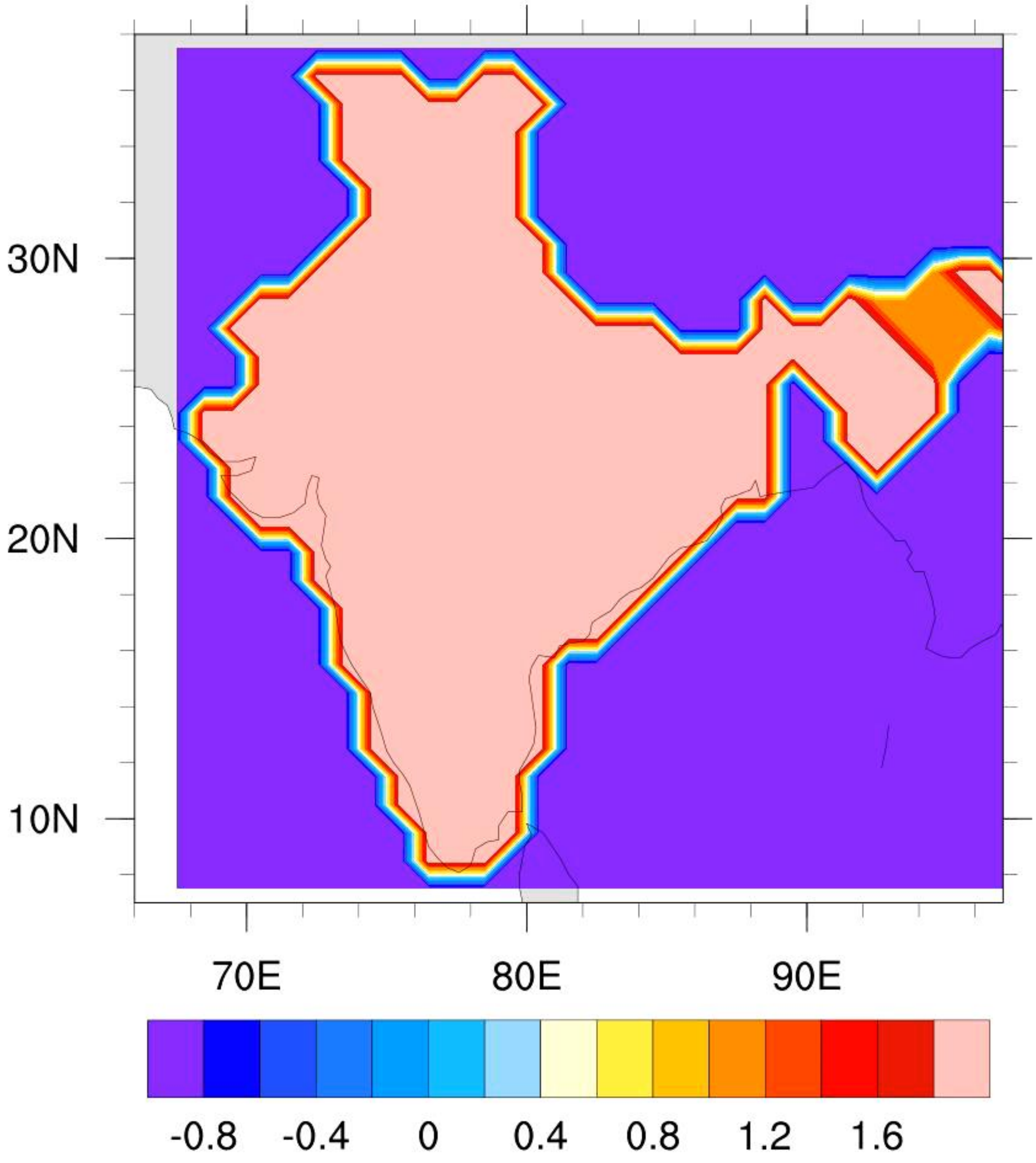
Markov Random Field Model

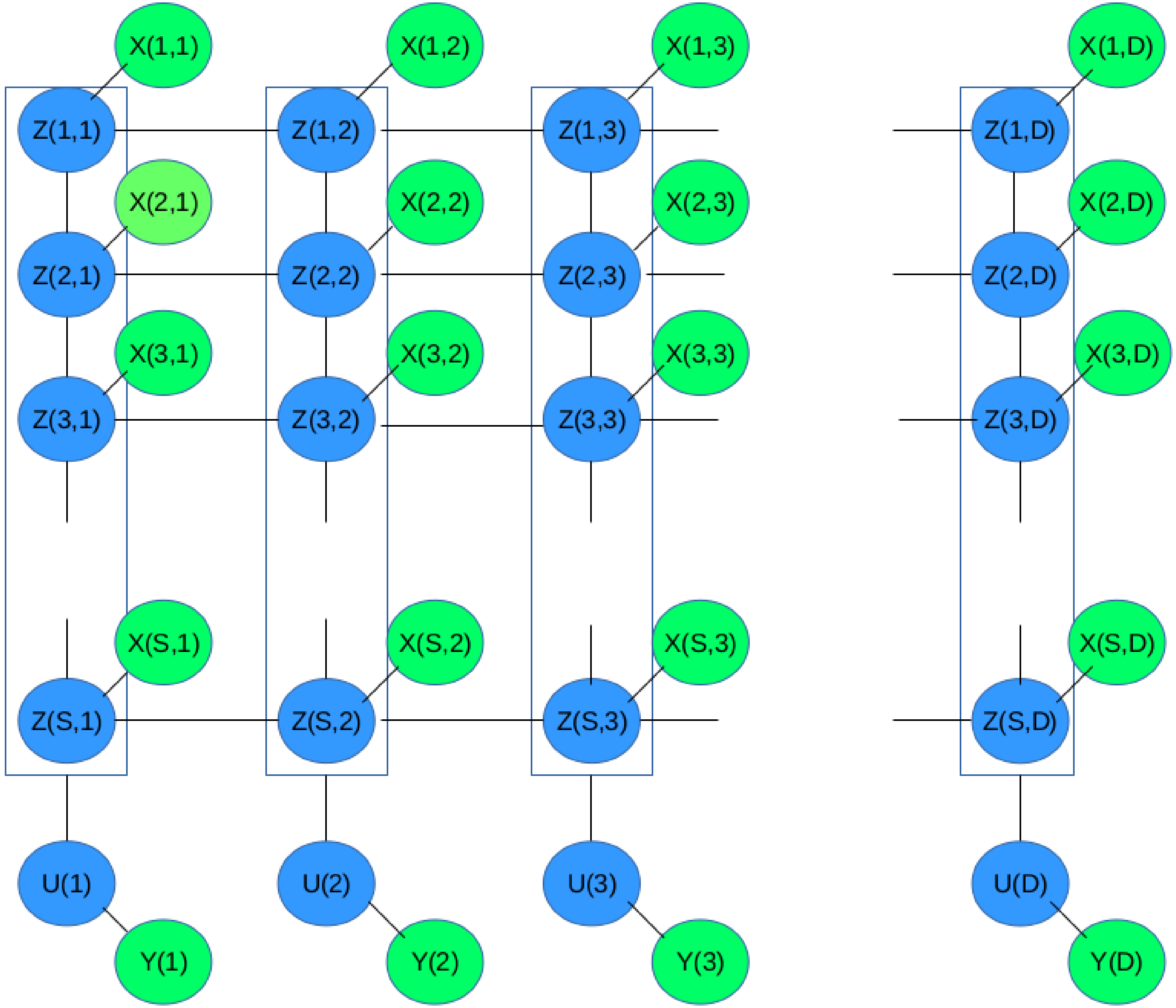
- Ψ (Z(s,t), Z(s',t)) = HIGH if Z(s,t)= Z(s',t)
  = LOW otherwise
- Ψ (Z(s,t), Z(s,t')) = HIGH if Z(s,t)= Z(s,t')
  = LOW otherwise
- Ensure spatio-temporal coherence of Z!
- Ψ (Z(s,t), X(s,t)) = Gamma (H(s,k), L(s,k))
  where k=Z(s,t)
- Ensures good fit of data

Inference Problem

- Z unknown, X known
- Find Z that maximizes joint distribution p(Z,X); conditioned on X
- Approximate Inference technique: Gibbs Sampling
- Iterative algorithm:

  - Sample each Z one by one, keeping others constant
  - Need to find conditional distribution of each Z
  - Easy due to certain properties of the joint distribution


70E
80E
90E
30N
20N
10N
-0.8
-0.4
0
0.4
0.8
1.2
1.6

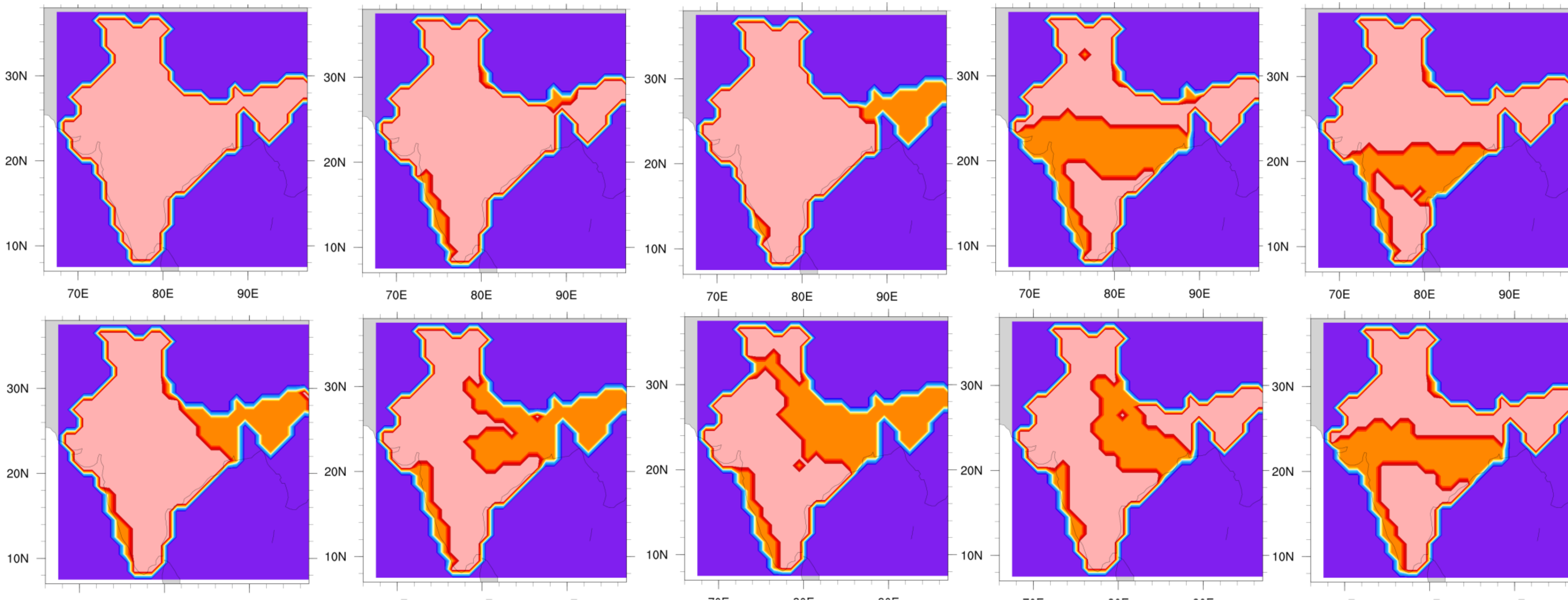

X(1,1)
X(1,2)
X(1,3)
X(1,D)
Z(1,1)
Z(1,2)
Z(1,3)
Z(1,D)
X(2,1)
X(2,2)
X(2,3)
X(2,D)
Z(2,1)
Z(2,2)
Z(2,3)
Z(2,D)
X(3,1)
X(3,2)
X(3,3)
X(3,D)
Z(3,1)
Z(3,2)
Z(3,3)
Z(3,D)
X(S,1)
X(S,2)
X(S,3)
X(S,D)
Z(S,1)
Z(S,2)
Z(S,3)
Z(S,D)
U(1)
U(2)
U(3)
U(D)
Y(1)
Y(2)
Y(3)
Y(D)

Binary Spatial Patterns

- Θ1, Θ2, Θ3.......: unknown S-dim binary vectors
- U(t) chooses one of these vectors for day t
- Z(t)-vector “similar” to ΘK where U(t)=K
-
- Ψ (Z(t), U(t)) = Hamming(Z(t),ΘK) where U(t)=K
- Ψ (Y(t), U(t)) = N(a(k), b(k)) where U(t)=K
-
- Gibbs Sampling extended to find U,Θ

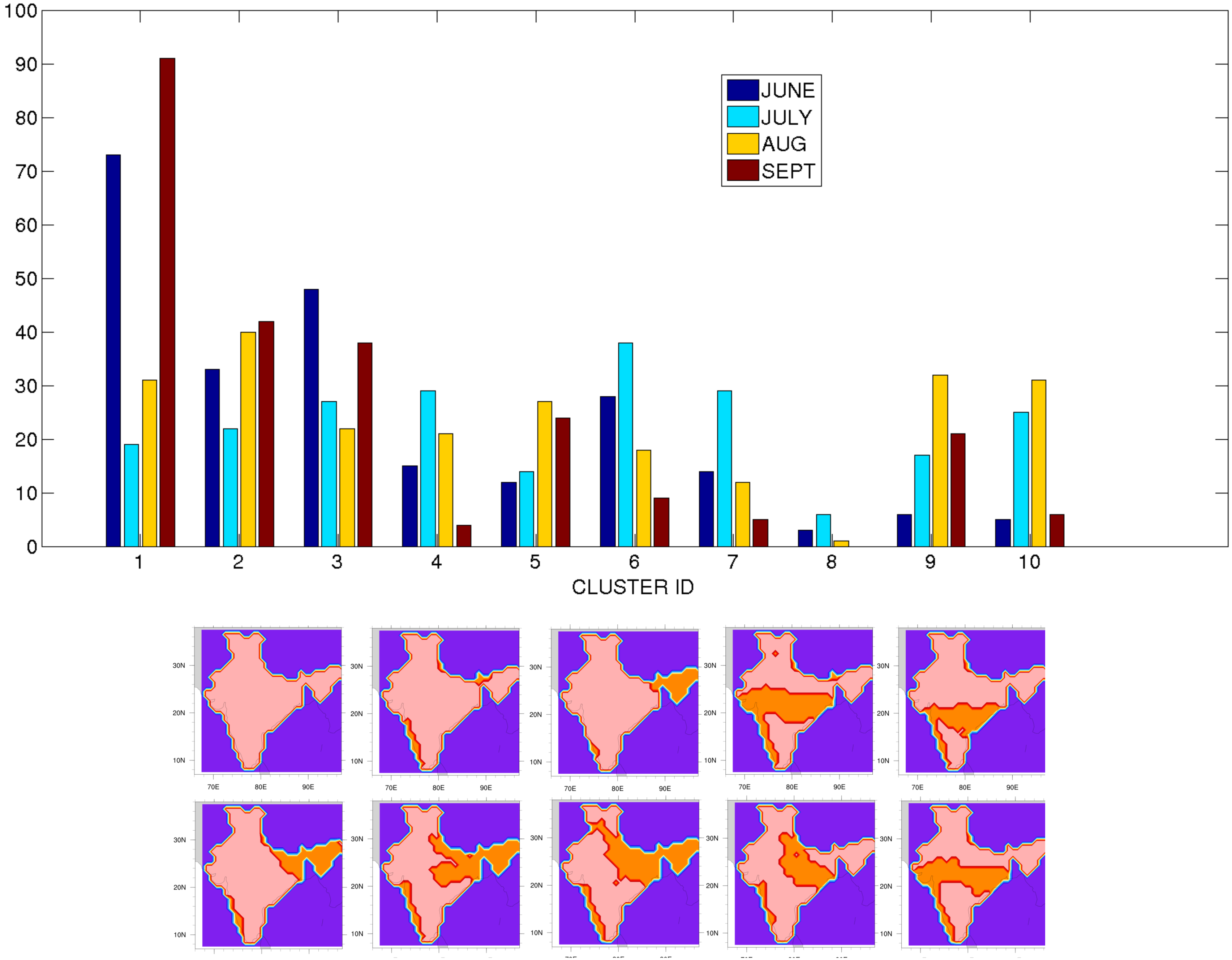
Binary Spatial Patterns

- Number of patterns = number of unique values taken by U
- Not fixed beforehand, known after inference
- We put an additional prior distribution on U:
- - discourages formation of many clusters
- - creates few large clusters
- - encourages each cluster to have members from different years

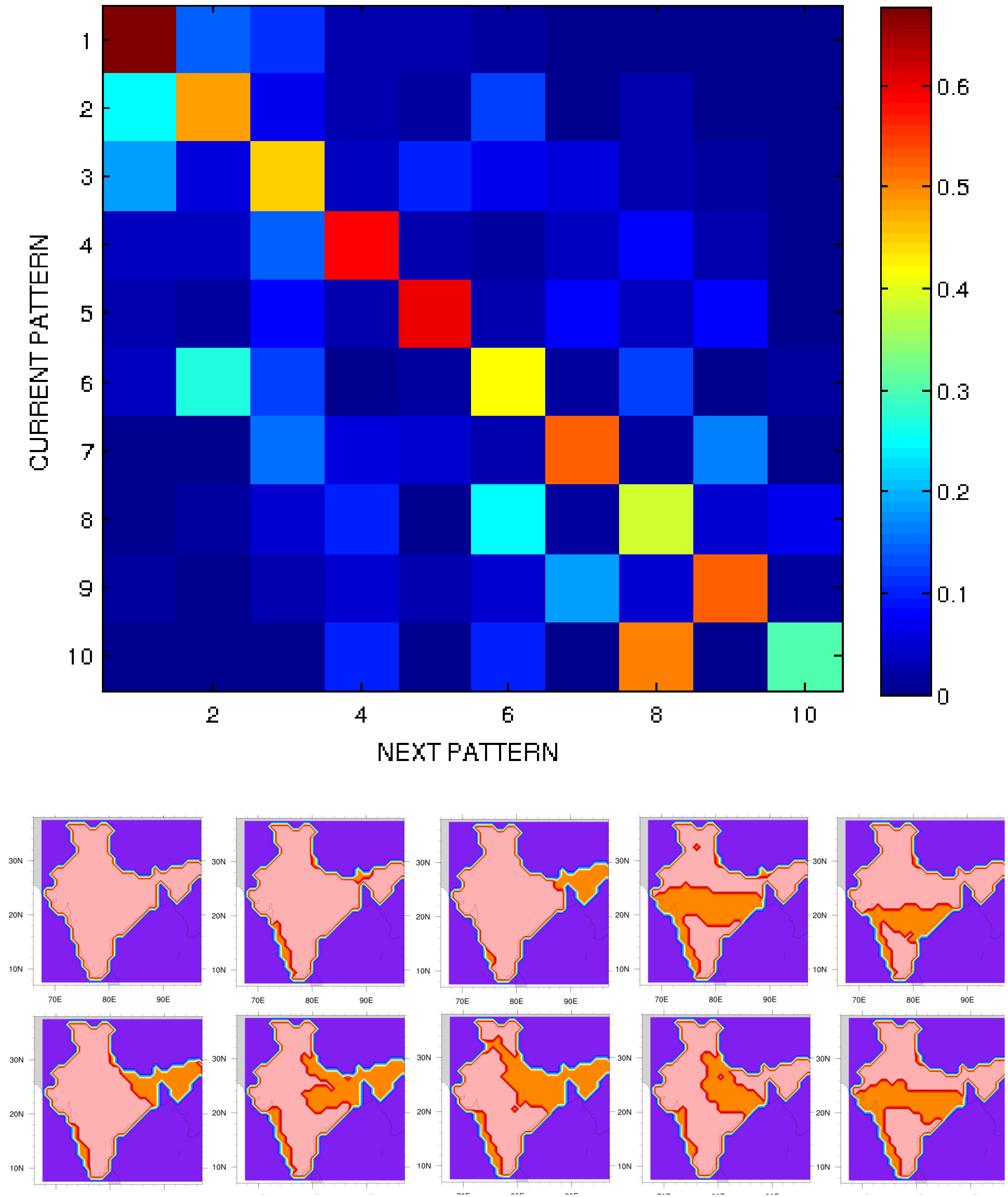

30N
20N
10N
70E
80E
90E

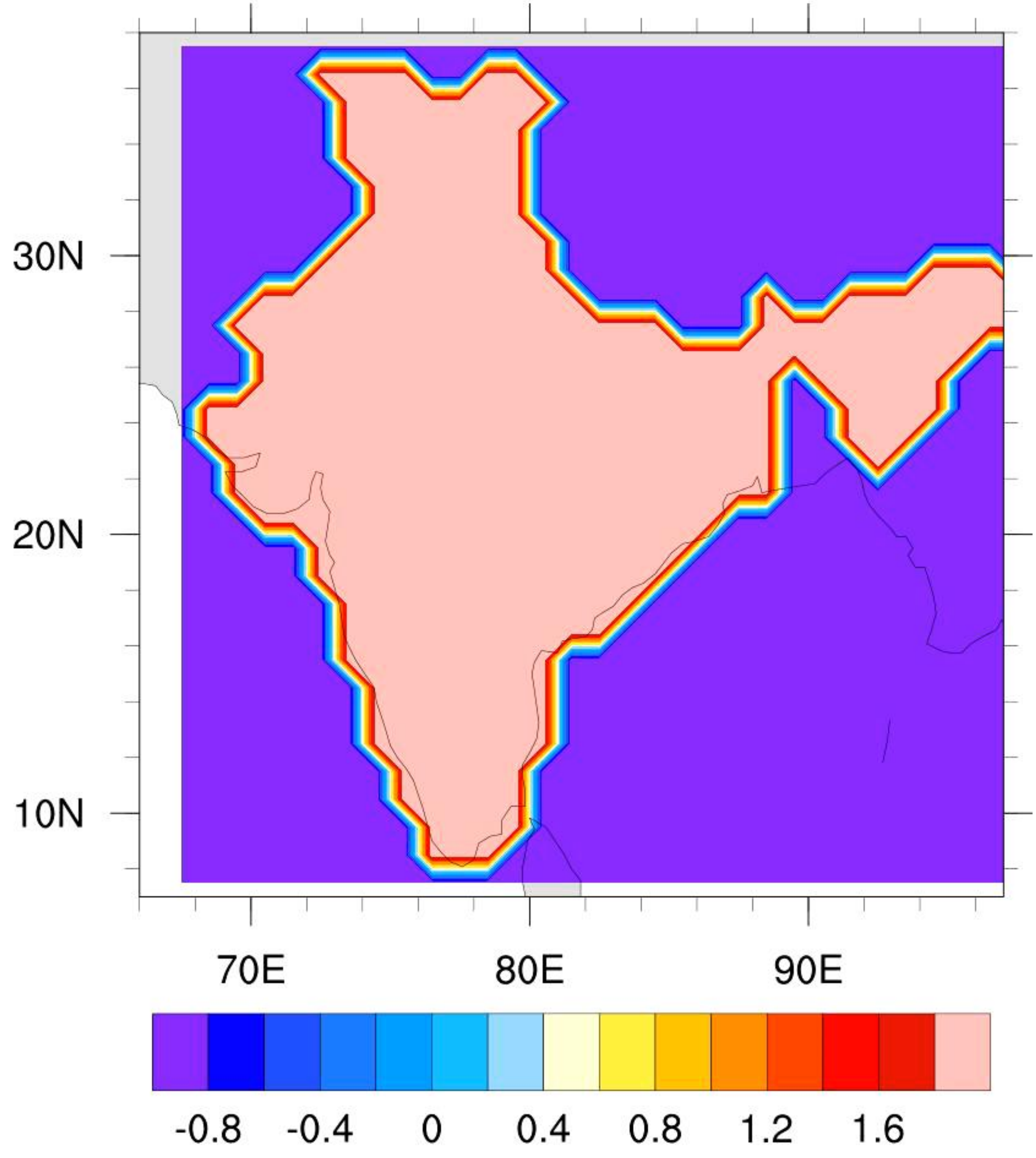
Binary Spatial Patterns

- Very coherent and interpretable
- Only 10 patterns but can represent 95% days of monsoon
- No need to specify number of clusters/patterns
- Very robust to time-period
- Open to study of dynamics/temporal relations



JUNE
JULY
AUG
SEPT
0
10
20
30
40
50
60
70
80
90
100
1
2
3
4
5
6
7
8
9
10
CLUSTER ID
30N
20N
10N
70E
80E
90E


CURRENT PATTERN
NEXT PATTERN
1
2
3
4
5
6
7
8
9
10
2
4
6
8
10
0.6
0.5
0.4
0.3
0.2
0.1
0
30N
20N
10N
70E
80E
90E


30N
20N
10N
70E
80E
90E
-0.8
-0.4
0
0.4
0.8
1.2
1.6

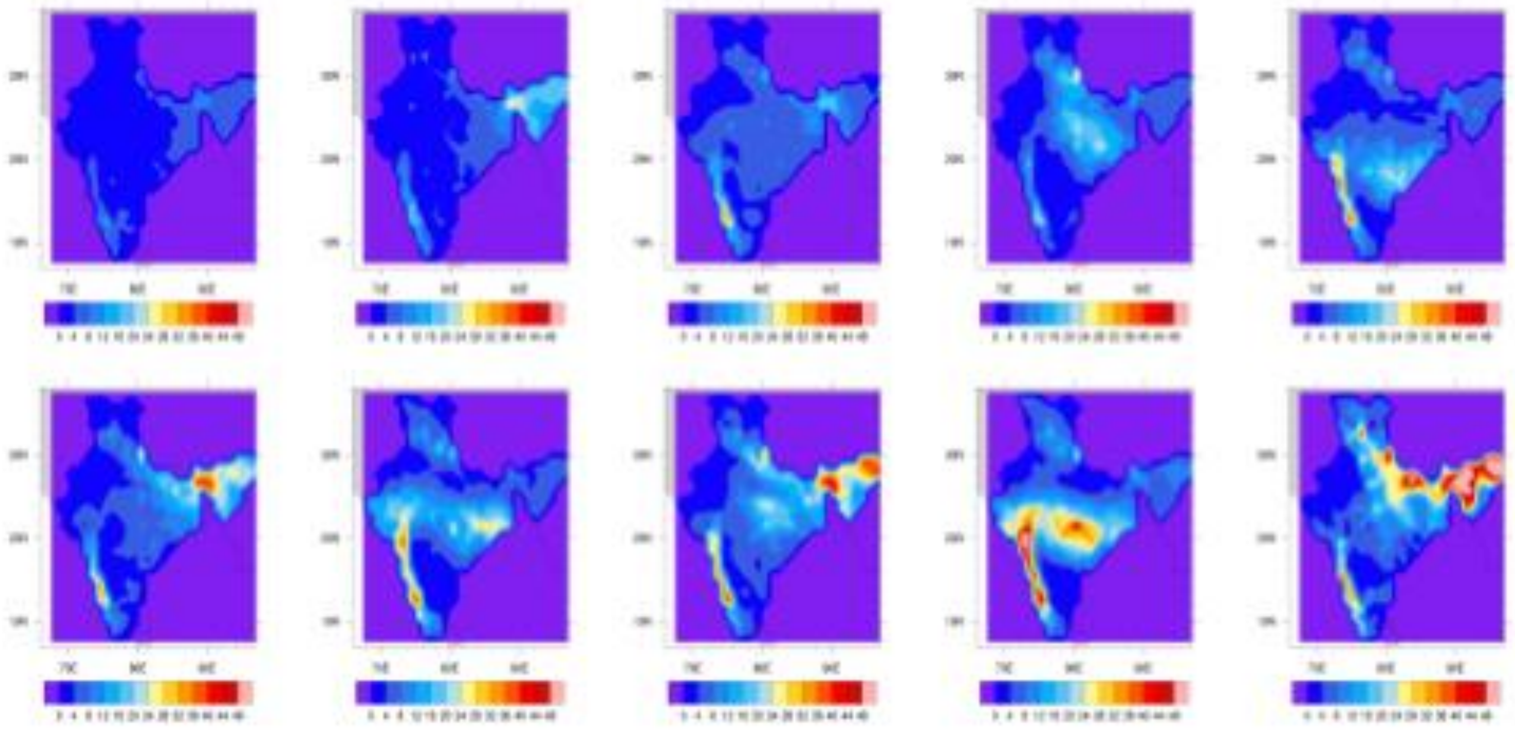


Figure 3: Canonical Rainfall Patterns (CRP) corresponding to the 10 CDPs shown above

Next Steps

- Look for similar patterns in other climatic fields like zonal winds and cloud cover
- Build multi-variable model
- Study the entire south Asian region
- Explore predictive power based on the transition patterns